CHAPTER 2

ASSESSMENT OF PRACTICAL SKILLS

2.1 Aim

The aim of the work presented in this section of the thesis is to study the reliability and validity of assessment.

Assessment of a manipulative skill of a pupil carried out by a number of judges using a check-list is ideally expected to be the same. However, one may ask, in reality, if that is so! The nature of agreement/disagreement among a number of judges is worth considering in order to answer such questions and to understand the reliability of assessment.

During the assessment of a set of pupils by a teacher, as is usually the case in real life, more than one skill may have to be assessed. One may ask if it would be less reliable when a judge concentrates on assessing more than one skill, instead of assessing only one skill. Here attempts have been made to find answers to such a question.

During a practical session, pupils carry out manipulative skills in order to collect data/record observations/deduce inferences which are submitted as written reports. As the task of assessing the skills by on-the-spot observation during practical sessions (i.e. Method -T) is time-consuming and thus difficult for a teacher, compared with the assessment carried out by checking the written report (i.e. Method -P) which can be done at home, it may be worthwhile to see if there is a significant correlation between the standard of achievement of a pupil assessed in both ways. So a comparison between these two methods of assessment is also undertaken. This will help us to find out if the result of assessment of a pupil by Method -P can be used as a measure of their performance in executing the manipulative skills during a practical session.

2.2 Development of the study materials - Preliminary runs

i) Method

Attempts were made to design an experiment which will allow us to (a) assess on the spot by observation of the performance of pupils in carrying out the skill of measuring a certain volume of liquid using a measuring cylinder, (b) assess by marking the final results of the whole experiment, and (c) study the reliability and validity of assessment of the skill.

Having given some thought, it was decided that the following experiment "reaction of Mg with dil. HCl" would be a simple experiment suitable for this purpose and would satisfy the scope and limitations described earlier.

Prior to the final run, a trial run was carried out by using about 25 pupils of an age range of 13-14 years from Norwich School, The Close, Norwich. The objective of this trial run was to find out the following:

(a) If the work-sheet is self-explanatory and easy to follow by all the pupils.

(b) Does the method work or is there any need for further modification?

(c) How many runs (number of experiments) can be carried out within a definite time limit (say one hour)?

(d) Are the apparatus and chemicals adequate for the work?

(e) How many pupils can be assessed by each judge for the measurement of volume and pouring out the liquid?

This trial run helped me to eliminate any deficiency in the preparation of the instruction in the work-sheet and also the arrangement of apparatus and chemicals.

(ii) Discussion of the out-come of the trial runs

The length of the Mg ribbon was a problem. As it was 5 cm in length, when it was dropped into the flask containing dil. HCl, occasionally it fell horizontally and sank into the acid. However most of the time it dropped vertically and a part of it remained above the acid. So it was decided to cut the Mg ribbon into 2 parts, each of 2.5 cm length and the pupils were asked to use 2 pieces of Mg ribbon for each experiment. In this way the ribbon dropped horizontally into the flask and sank into the acid each time. No other modification was necessary.

2.3 Final run

(i) Method

The tasks for the pupils were to find out the time taken by 2 pieces (2.5 cm each piece) of Mg ribbon to dissolve completely in a mixture of dil. HCl and water (see work-sheet 2.1 in Appendix 1.1).

Four sets of 4th form pupils (14-15 year old) from a local comprehensive school (Earlham School) took part in this experiment. About 20 pupils (boys and girls) were in each set and they were randomly allocated a work-station before the work-session.

One large teaching laboratory in the School of Chemical Sciences at the University of East Anglia was used for this purpose. Each work-station was numbered so that each pupil, randomised at the door and provided with numbered labels, could come quickly to the matching numbered desk.

In order to facilitate the on-the-spot assessment by the judges who each carried a check-list and identified the pupils only by their numbers, the following arrangements were made. The pupils carried identifying number labels in both of their arms and the number plates at the work-stations were raised to a reasonable height so that the judges could see it clearly from a reasonable distance.

Each session lasted for about an hour. In between two sessions, there was an interval of about 15-30 minutes during which the used glass-ware was removed and new chemical and apparatus were laid out. A list of chemicals and apparatus used in this experiment is given in Appendix 1.2.

Each pupil was asked to carry out 6 runs (experiments) using 6 acid-water mixtures of different volume ratios. Work-sheets and other required materials were supplied at the work-desk before the start of the experiment.

Nine qualified science-teachers from local schools participated as judges. They were divided into 3 panels and were assigned for carrying out the on-the-spot assessment by observation of the performance of the skill in the following way:

Work-scheme for the judges

Session	Panel A	Panel B	Panel C
0 (trial run)	M + P	M + P	M + P
1	M	P	M + P
2	M + P	M	P
3	P	M + P	M

M and P stand for measuring volume by a cylinder and pouring the liquid out of the cylinder respectively.

The experiment carried out by the first set of pupils was used by the judges as a trial run to gain a clear idea of the nature of their task in this experiment. This was helpful for their asessment work in the three subsequent sessions.

Before the start of the experiment, the judges were briefed about the nature of the task and clear instructions were given about the job in each session, as has been outlined in the table above.

It was thought that, among all the skills involved in this experiment, the two most important skills to affect the final results, i.e. the time for Mg to dissolve, are: (1) measuring the volume of acid/acid + water mixture, and (2) pouring out the liquid into the flask. Each judge assessed each pupil once for measurement of volume (M) or pouring out the liquid (P) or both (M) and (P), according to the work-scheme given earlier. The judges were instructed to give "1" mark for correctly executing the skill and "0" for failing to do so.

The judges checked the accuracy of the volume measurement by taking a close look at the measuring cylinder and seeing if the lower meniscus was within

± 0.5 cm^3 of the volume to be measured. The accuracy of the pouring skill was checked by looking at each pupil's effort to drain the last drop of liquid out of the cylinder.

In order to assess on the basis of the outcome of the extended task, the results of the experiment submitted by the pupils in the form of a written report (i.e. the time taken for Mg to dissolve reported by the pupil) were checked against the values determined by me. For this purpose each of the 6 experiments [see the work-sheet 2.1 in Appendix 1.1] was repeated 8 times by me. The volume of acid was measured accurately, the pouring out operation was carried with great care and finally the exact time taken by the Mg ribbon to dissolve completely was recorded accurately. The mean value of time for each experiment (with the standard deviation as the accepted error limit) was used as the standard result to carry out this assessment. A table containing these results and s.d. values is given in the Appendix 1.3. Probable explanation for the variation of the s.d values with the variation of acid:water composition is also given there.

(ii) Results and Discussion

As the first assessment session during the experiment was used as a trial run and referred to as session -0 in this section, results of only the 2nd, 3rd and 4th sessions of the experiment will be used in the calculation here and they will be referred to hereafter as Session -1, Session -2 and Session -3 respectively.

According to the marking scheme, each pupil was marked in the following way: a panel of 3 judges marked only the volume measurement skill, another panel of 3 judges marked only the pouring out skill and the 3rd panel of 3 judges marked both the skills. As each judge was instructed to give 1 mark for correct execution of a skill, each pupil could get a maximum of 6 marks for volume measurement, and 6 marks for pouring out the liquid.

The data in the judges' assessment sheets for all the 3 sessions (see Appendix 1.4) has been analysed and the information available in these tables was used to study the reliability and validity of assessment.

In order to look at the reliability of assessment, the following two aspects have been considered:
(i) comparison has been made between the achievement of the pupils for their volume measurement skill assessed by a panel (say "A") of 3 judges for the volume measurement alone, and that assessed by another panel (say "C") of 3 judges for both the volume measurement and pouring out the liquid. Similar comparison can be made between the achievement of the pupils for their liquid pouring skill assessed by a panel (say "B") of 3 judges for the pouring out liquid alone, and that assessed by the panel "C" for both volume measurement and pouring out liquid mentioned earlier.

(ii) The possible ways in which a panel of 3 judges can mark a pupil during assessing a skill are shown in the table below:

Possibilities of marks for a pupil from 3 judges			Overall achievement of the pupil	Agreement among the judges (Yes: 1, No:0)
1	1	1	3 marks	1
1	1	0	2 marks	0
1	0	0	1 mark	0
0	0	0	0 mark	1

When a pupil gets 3 marks (1 + 1 + 1) or "0" mark (0 + 0 + 0), the 3 judges in a panel are in perfect agreement, otherwise the judges are not in perfect agreement. A situation of perfect agreement among the judges is represented by "1", while the lack of perfect agreement is indicated by "0".

Tables 2.1, 2.2, 2.3 and 2.4 contain summaries of the analysis of the data related to the reliability of the assessment studies. Tables 2.1 and 2.2, derived from further analysis of data in Table 2.6, 2.7 and 2.8, show the mean values of achievement when all the pupils in a session have been assessed for one skill, e.g. volume measurement/pouring out liquid, by a panel of judges, compared with the situation when both the skills have been assessed by another panel of judges.

Tables 2.3 and 2.4, derived from further analysis of data in Table 2.9,2.10

2.11, show the mean values of total agreement among the judges in a panel when all the pupils in a session have been assessed for one skill, e.g. measuring volume/pouring out liquid, by a panel of judges, compared with the situation when both the skills have been assessed by another panel of judges.

It appears that the mean value of the achievement for any set (or, for all the 3 sets together) for volume measurement does not differ significantly whether a panel of judges assessed it alone or another panel of judges assessed both skills of measuring volume and pouring a liquid.

A similar conclusion can be drawn for the mean value of the achievement for pouring out skill, whether a panel of judges assessed that skill alone or another panel of judges assessed it together with the volume measurement skill.

The mean value of performance for the pouring out skill is much higher than that for the volume measurement skill. The results suggest that the pupils in every set have been able to carry out this "pouring out the liquid" skill very effectively, almost perfectly. However poor performance (about 33% of the total marks) for volume measurement suggests that the pupils would require further lessons in this skill. The final results of the experiments, which have been used as the basis for assessment by Method -P, are directly influenced by the pupils' ability to carry out the volume measurement skill efficiently.

With regard to the agreement/disagreement among the judges, it can be said from the Tables 2.3 and 2.4 that the judges are in total agreement for about 50% of the sample during the assessment of the volume measurement skill, irrespective of whether it was assessed alone or assessed in conjunction with the pouring out skill. The amount of total agreement among the judges for assessing pouring out skill was about 100%. As with the volume measurement skill, the pouring out skill showed no significant difference in the amount of total agreement depending on whether it was assessed alone or assessed with the volume measurement skill.

For further analysis of the data of total agreement among the judges, several 2X3 ANOVA based on randomised groups were carried out. They were as follows:

(i) 2 ways of assessing volume measurement skill (A) X 3 panels of judges (B)

(ii) 2 ways of assessing pouring out skill (A) X 3 panels of judges (B)

(ii) 2 ways of assessing volume measurement skill (A) X 3 sessions (sets of pupils)(B)

(iv) 2 ways of assessing pouring out skill (A) X 3 sessions (sets of pupils)(B)

Information derived from the ANOVA is in Table 2.5.

None of the F_A value is significant even at $p < 0.05$ level. This suggests that whichever way a skill, i.e. measurement of volume or pouring out liquid, is assessed, the amount of total agreement among the judges is unaffected.

This finding lends support to the earlier conclusion regarding the reliability of assessment based on the fact that the number of total agreements does not differ significantly between the two ways of assessing volume measurement or pouring out skill.

Furthermore, as none of the F_B or F_{AXB} is significant even at $p < 0.05$, assessment being carried out by three panels of judges and in three sessions also do not have any significant effect on the number of agreement among the judges and hence on the reliability of the assessment.

The achievement based on the assessment by Method -T of performing each of the practical skills for all the pupils in each of the 3 sets working in 3 sessions has been presented in Tables 2.6, 2.7 and 2.8. Appendix 1.4 contains the "raw-data", i.e. 3 panels (A, B, C) of judges' mark-sheets for all the 3 assessment sessions. The data in each session were analysed in the following way. Marks obtained by each pupil for his/her volume measurement/liquid pouring out skill were added togehter and presented as the achievement data based on the assessment Method -T in Tables 2.6, 2.7, 2.8. As each pupil was assessed by 3 judges for volume measurement alone, 3 judges for pouring out alone and 3 judges for both skills, the maximum marks that a pupil can get for correct execution of each skill is 6. In each table, there is also a column of data showing the achievement of these pupils based on the assessment by Method -P which involves the judgement of the accuracy of the final results of the experiment (i.e. the time taken by the Mg ribbon to dissolve completely). A maximum of 6 marks can be obtained by a pupil for the correct results of 6 runs in experiment 2.1.

Using judges' mark-sheets (see Appendix 1.4)/Pupils' achievement data in Tables 2.6, 2.7, and 2.8, situations of agreement/disagreement among the judges in assessing the pupils in all 3 sessions have been analysed and are listed in Tables 2.9, 2.10 and 2.11.

The marks awarded by the judges (assessment by method-T) for volume measurement and pouring out a liquid was plotted against the marks, based on the assessment by Methid -P, for all the pupils in the 3 sessions in Figures 2.1 and 2.2 respectively. Such a comparison can be used as a measure of validity of assessment. The data plotted in each figure appears to be widely scattered and it is quite difficult to distinguish any clear relationship (trend) between data gathered from two methods of assessment. For this Pearson-r (i.e. Pearson product moment correlation coefficient) values were calculated and the results are shown in Table 2.12. The r-values appear to lie within ± 0.5 suggesting the lack of a very strong (near to perfect) correlation. However, it is worthwhile to look at the statistical significance of the existing correlation as shown by the r-value.

Here the experimental hypothesis is one tailed and a positive correlation is expected. It is thought that a pupil's achievement for performing volume measurement skill/pouring out skill assessed by Method -T will be associated with the achievement of the pupil assessed by Method -P (i.e. marking the written report on the outcome of the extended task of the whole experiment). According to the "Null Hypothesis" the correlation "γ" value obtained from calculation could be due to chance factor. So the calculated "γ" value is compared with the critical value of "γ" at the corresponding degree of freedom df = 58 (df = N-2, where N = no. of subjects in the experiment).

One of the "γ" values calculated for the volume measurement skill $\gamma = 0.306$ (for the 6th run) appears to be higher than the critical value at $p = 0.025$ but lower than the critical value at $p = 0.01$. This suggests that the calculated "γ" value is significant at $p < 0.025$ level, i.e. % probability of correlation by chance factor is equal to or less than 2.5% and thus the "Null Hypothesis" is rejected and experimental hypothesis is accepted at $p < 0.025$ level. However, the rest of the "γ" values calculated for volume measurement skill and all the "γ" values for pouring out skill seem to be lower than the lowest possible critical value of γ (= 0.2306) for df = 58 which is at $p = 0.05$. This suggests that the % probability of the correlation occurring by chance factor is higher than 5% and thus the "γ" values are not significant statistically.

A significant correlation, even though at $p < 0.025$ level, shown (by the result of one of the runs of volume measurement experiment) between the achievement data of the pupils when assessed by method -T and that when assessed by method -P, suggests that the accuracy of the volume measurements do indeed have a strong influence on the final result, i.e. time needed for Mg ribbon to dissolve completely and assessment of the final result submitted as a written report can be used as a substitute for the assessment of volume measurement skill on the spot.

Although the "pouring out liquid" skill is an important technique in the whole experiment and should have a strong influence on the final result, the lack of a strong correlation between the achievement data from method -T and that from method -P is likely to be due to the fact that performance of both the volume measurement skill and pouring out skill affect the accuracy of the final result and hence the pupils' achievement in assessment by method -p. From the discussion of the experimental results, which has been given earlier, it has become clear that pupils perform the pouring out skill almost perfectly, while their performance in volume measurement skill is not very good (obtained only 33% of the total marks). So, pupils' low achievement data obtained from the assessment method -p is perhaps the direct consequence of their poor performance in volume measurement, while the contribution of the pouring out skill on the final result can be considered as constant (all the pupils performed it very well).

Table 2.1

Achievement of pupils in carrying out 2 manipulative skills involved in experiment 2.1

(The data-analyses were based on the fact that the 3 sets of pupils were assessed in 3 sessions)

Session		Skill of measuring volume of liquid		Skill of pouring out liquid	
		M	M+P	P	M + P
1	x	1.05	1.30	2.85	2.80
	s.d.	±0.86	±0.90	±0.36	±0.40
2	x	1.00	0.95	3.00	2.85
	s.d.	±1.09	±0.92	±0.00	±0.36
3	x	0.80	0.90	2.75	2.75
	s.d.	±0.93	±0.94	±0.43	±0.43
1 + 2 + 3	x	0.95	1.05	2.87	2.80
	s.d.	±0.97	±0.94	±0.34	±0.40

x = mean value of achievement per pupil (maximum value is 3)
s.d. = standard deviation in "x"
M/P = only volume measurement/pouring out skill was assessed by a panel of judges in a session
M + P/M + P = Both the skills were assessed by a panel in a session (only the underlined skill assessment was used in calculation)

Table 2.2

Achievement of pupils in carrying out 2 manipulative skills involved in experiment 2.1
(The data analyses were based on the fact that 3 panels of judges carried out the assessment)

Panel of judges		Skill of measuring volume of liquid		Skill of pouring out liquid	
		M	M + P	P	M + P
A	x	1.05	0.95	2.75	2.85
	s.d.	±0.86	±0.92	±0.43	±0.36
B	x	1.00	0.90	2.85	2.75
	s.d.	±1.09	±0.94	±0.36	±0.43
C	x	0.80	1.30	3.00	2.80
	s.d.	±0.93	±0.90	±0.0	±0.40
A + B + C	x	0.95	1.05	2.87	2.80
	s.d.	±0.97	±0.94	±0.34	±0.40

x = mean value of achievement per pupil (maximum value is 3)
s.d. = standard deviation in "x"
M/P = only volume measurement/pouring out skill was assessed by a panel of judges in a session
M + P/M + P = Both the skills were assessed by a panel of judges in a session (only the underlined skill assessment was used in calculation)

TABLE 2.3

Amount of total agreement among the 3 judges in a panel in assessing the pupils' skill in experiment 2.1
(The data analyses were based on the fact that the 3 sets of pupils were assessed in 3 sessions)

Session		Skill of measuring volume of liquid		Skill of pouring out liquid	
		M	M + P	P	M + P
1	x	0.35	0.30	0.85	0.85
	s.d.	±0.48	±0.46	±0.36	±0.40
2	x	0.60	0.45	1.00	0.85
	s.d.	±0.49	±0.50	±0.00	±0.36
3	x	0.50	0.50	0.75	0.75
	s.d.	±0.50	±0.50	±0.43	±0.43
1 + 2 + 3	x	0.50	0.42	0.87	0.80
	s.d.	±0.50	±0.49	±0.34	±0.40

x = mean value of agreement (maximum value is 1)
s.d. = standard deviation in "x"
M/P = only volume measurement/pouring out skill was assessed by a panel of judges in a session.
M + P/M + P = Both skills were assessed by a panel of judges in a session (only the underlined skill assessment was used in calculation)

Table 2.4

Amount of total agreement among the 3 judges in a panel in assessing the pupils' skill in experiment 2.1
(The data analyses were based on the fact that 3 panels of judges carried out the assessment)

Panel of judges		Skill of measuring volume of liquid		Skill of pouring out liquid	
		M	M + P	P	M + P
A	x	0.35	0.45	0.75	0.85
	s.d.	±0.48	±0.50	±0.43	±0.36
B	x	0.60	0.50	0.85	0.75
	s.d.	±0.49	±0.50	±0.36	±0.43
C	x	0.55	0.30	1.00	0.85
	s.d.	±0.50	±0.46	±0.36	±0.40
A + B + C	x	0.50	0.42	0.87	0.80
	s.d.	±0.50	±0.49	±0.34	±0.40

x = mean value of agreement (maximum value is 1)
s.d. = standard deviation in "x"
M/P = only volume measurement/pouring out skill was assessed by a panel of judges in a session
M + P/M + P = Both skills were assessed by a panel of judges in a session. (only the underlined skill assessment was used in calculation)

Table 2.5

Summary tables for several 2 x 3 ANOVA based on completely randomised groups for the reliability of assessment (agreement among the judges) of pupils' ability to perform scientific manipulative skill in experiment 2.1

(a) Measuring volume of liquid using measuring cylinder

Source of variance	SS	df	MS	F (obsv.)	F (crit) at p=0.05	Significance of F (obsv.)
A (2 method gr.)	0.209	1	0.209	0.837	4.00	n.s. ($p < 0.05$)
B (3 panels of judges)	0.517	2	0.259	1.040	3.15	n.s. ($p < 0.05$)
A × B	0.616	2	0.308	1.234	3.15	n.s. ($p < 0.05$)
Error	28.450	114	0.249	-	-	
Total	29.792	119	-	-	-	

(b) Pouring liquid out of measuring cylinder

Source of variance	SS	df	MS	F (obsv.)	F (crit) at p=0.05	Significance of F (obsv.)
A (2 method gr.)	0.133	1	0.133	0.957	4.00	n.s. ($p < 0.05$)
B (3 panels of judges)	0.267	2	0.134	0.960	3.15	n.s. ($p < 0.05$)
A × B	0.467	2	0.234	1.680	3.15	n.s. ($p < 0.05$)
Error	15.800	114	0.138	-	-	
Total	16.667	119	-	-	-	

/continued

Table 2.5 (continued)

(c) Measuring volume of liquid using measuring cylinder

Source of variance	SS	df	MS	F (obsv.)	F (crit) at p=0.05	Significance of F (obsv.)
A (2 method gr.)	0.208	1	0.208	0.833	4.00	n.s. ($p < 0.05$)
B (3 sessions)	1.067	2	0.533	2.138	3.15	n.s. ($p < 0.05$)
A × B	0.067	2	0.033	0.134	3.15	n.s. ($p < 0.05$)
Error	28.450	114	0.249	-	-	
Total	29.792	119	-	-	-	

(d) Pouring liquid out of measuring cylinder

Source of variance	SS	df	MS	F (obsv.)	F (crit) at p=0.05	Significance of F (obsv.)
A (2 method gr.)	0.133	1	0.133	0.957	4.00	n.s. ($p < 0.05$)
B (3 sessions)	0.617	2	0.308	2.219	3.15	n.s. ($p < 0.05$)
A × B	0.117	2	0.058	0.421	3.15	n.s. ($p < 0.05$)
Error	15.800	114	0.138	-	-	
Total	16.667	119	-	-	-	

In Tables 2.5 (a) - (d):

F (obsv.)	denotes the value of F ratio obtained from the experimental results.
F (crit)	denotes the critical value of F ratio at a given level of significance (i.e. p value) and at the appropriate df values (see reference 50 for the relevant statistical tables).
n.s.	means "not significant". This refers to the fact that the value of observed F ratio is less than the critical value of F at p=0.05 and thus F (observed) is not significant at $p < 0.05$ level.

Table 2.6

Achievement of pupils (participating in Session 1) for carrying out 2 manipulative skills (measuring volume of liquid and pouring out liquid) involved in experiment 2.1, assessed both by (1) Method -T (i.e. on the-spot observation of the performance of the techniques) and (2) Method -P (i.e. checking the reported results/final product of the experiment)

Pupil No.	Achievement in assessment by Method -T						Achievement in assessment by Method -P
	Marks for measuring volume			Marks for pouring liquid			Marks for reported results
	M	M+P	Total marks (6 max)	P	M+P	Total marks (6 max)	(6 max)
1	2	1	3	3	3	6	2
2	-	-	-	-	-	-	-
3	3	1	4	3	3	6	4
4	2	0	2	2	3	5	4
5	1	1	2	3	2	5	6
6	2	3	5	3	3	6	5
7	0	1	1	3	2	5	4
8	1	0	1	3	3	6	1
9	-	-	-	-	-	-	-
10	1	1	2	3	3	6	3
11	1	1	2	3	3	6	6
12	0	1	1	2	2	4	5
13	1	3	4	3	3	6	2
14	1	2	3	3	3	6	4
15	2	1	3	3	3	6	4
16	2	2	4	3	3	6	3
17	0	0	0	3	3	6	3
18	0	3	3	2	2	4	5
19	1	1	2	3	3	6	4
20	1	2	3	3	3	6	4
21	0	1	1	3	3	6	3
22	-	-	-	-	-	-	-
23	0	1	1	3	3	6	1

M/P: Only volume measurement/pouring out skill was assessed (max. marks for each skill = 3)

M + P/M + P: Both the skills were assessed, but only the underlined skill's mark is involved in the calculation.

Table 2.7

Achievement of pupils (participating in Session 2) for carrying out 2 manipulative skills (measuring volume of liquid and pouring out liquid) involved in experiment 2.1, assessed both by (1) Method -T (i.e. on-the-spot observation of the performance of the techniques) and (2) Method -P (i,e, checking the reported results/final product of the experiment).

	Achievement in assessment by Method -T						Achievement in assessment by Method -P
Pupil No.	Marks for measuring volume			Marks for pouring liquid			Marks for reported results
	M	M+P	Total marks (6 max)	P	M+P	Total marks (6 max)	(6 max)
1	2	1	3	3	3	6	3
2	1	0	1	3	3	6	6
3	0	0	0	3	3	6	0
4	0	1	1	3	3	6	4
5	0	0	0	3	2	5	4
6	0	1	1	3	3	6	5
7	1	1	2	3	3	6	4
8	1	1	2	3	3	6	1
9	0	0	0	3	3	6	3
10	0	0	0	3	3	6	6
11	1	0	1	3	3	6	3
12	2	1	3	3	3	6	4
13	1	1	2	3	3	6	5
14	3	2	5	3	2	5	3
15	2	2	4	3	3	6	5
16	0	1	1	3	3	6	5
17	3	3	6	3	3	6	4
18	3	3	6	3	3	6	2
19	0	1	1	3	3	6	4
20	0	0	0	3	2	5	3

M/P: Only volume measurement/pouring out skill was assessed (max. marks for each skill = 3)

M + P/M + P: Both the skills were assessed, but only the underlined skill's mark is involved in the calculation.

Table 2.8

Achievement of pupils (participating in Session 3) for carrying out 2 manipulative skills (measuring volume of liquid and pouring out liquid) involved in experiment 2.1, assessed both by (1) Method -T (i.e. on-the-spot observation of the performance of the techniques) and (2) Method -P (i.e. checking the reported results/final product of the experiment).

	Achievement in Assessment by Method -T						Achievement in assesment by Method -P
Pupil No.	Marks for measuring volume			Marks for pouring liquid			Marks for reported results
	M	M+P	Total marks (6 max)	P	M+P	Total marks (6 max)	(6 max)
1	0	1	1	3	3	6	4
2	0	1	1	3	3	6	4
3	0	0	0	3	2	5	1
4	1	3	4	3	3	6	5
5	0	0	0	2	2	4	5
6	1	0	1	3	3	6	6
7	0	0	0	2	2	4	4
8	2	2	4	2	2	4	2
9	2	2	4	2	3	5	4
10	2	2	4	3	3	6	5
11	3	2	5	3	3	6	5
12	0	1	1	3	3	6	6
13	1	0	1	3	3	6	2
14	1	1	2	3	3	6	4
15	2	2	4	3	3	6	3
16	0	0	0	3	3	6	0
17	1	0	1	3	2	5	5
18	0	1	1	2	3	5	3
19	0	0	0	3	3	6	4
20	0	0	0	3	3	6	0

M/P: Only volume measurement/pouring out skill was assessed (max. marks for each skill = 3)

M + P/M + P: Both the skills were assessed, but only the underlined skill's mark is involved in the calculation.

Table 2.9

Situation of total agreement ("1")/disagreement ("0") among 3 judges in a panel during the assessment in session 1.

Pupil No.	M by panel A	P by panel B	M + P by panel C	
1	0	1	0	1
2	–	–	–	–
3	1	1	0	1
4	0	0	1	1
5	0	1	0	0
6	0	1	1	1
7	1	1	0	0
8	0	1	1	1
9	–	–	–	–
10	0	1	0	1
11	0	1	0	1
12	1	0	0	0
13	0	1	1	1
14	0	1	0	1
15	0	1	0	1
16	0	1	0	1
17	1	1	1	1
18	1	0	1	0
19	0	1	0	1
20	0	1	0	1
21	1	1	0	1
22	–	–	–	–
23	1	1	0	1

M : only "Measurement of volume" skill assessed

P : only "Pouring out liquid" skill assessed

M + P : Both the skills assessed

Table 2.10

Situation of total agreement ("1")/disagreement ("0") among 3 judges in a panel during the assessment in session 2.

Pupil No.	M by panel B	P by panel C	M + P by panel A	
1	0	1	0	1
2	0	1	1	1
3	1	1	1	1
4	1	1	0	1
5	1	1	1	0
6	1	1	0	1
7	0	1	0	1
8	0	1	0	1
9	1	1	1	1
10	1	1	1	1
11	0	1	1	1
12	0	1	0	1
13	0	1	0	1
14	1	1	0	0
15	0	1	0	1
16	1	1	0	1
17	1	1	1	1
18	1	1	1	1
19	1	1	0	1
20	1	1	1	1

M : only "Measurement of volume" skill assessed

P : only "Pouring out liquid" skill assessed

M + P : Both the skills assessed

Table 2.11

Situation of total agreement ("1")/disagreement ("0") among 3 judges in a panel during the assessment in session 3.

Pupil No.	M by panel C	P by panel A	M + P by panel B	
1	1	1	0	1
2	1	1	0	1
3	1	1	1	0
4	0	1	1	1
5	1	0	1	0
6	0	1	1	1
7	1	0	1	0
8	0	0	0	0
9	0	0	0	1
10	0	1	0	1
11	1	1	0	1
12	1	1	0	1
13	0	1	1	1
14	0	1	0	1
15	0	1	0	1
16	1	1	1	1
17	0	1	1	0
18	1	0	0	1
19	1	1	1	1
20	1	1	1	1

M : only "Measurement of volume" skill assessed

P : only "Pouring out liquid" skill assessed

M + P : Both the skills assessed

Table 2.12

Pearson-r (Pearson product moment correlation coefficient) between the achievements of pupils after being assessed by observation of their performance (Method -T) and that after being assessed by checking the reported results of the experiment (Method -P) for two practical skills (a) measurement of volume of liquid and (b) pouring out liquid, in the experiment 2.1

No. of run of the experiment	Measurement of volume of liquid	Pouring out liquid
1st	0.159	0.078
2nd	-0.160	-0.011
3rd	0.223	0.111
4th	-0.023	-0.044
5th	0.100	-0.183
6th	0.306	0.021
Combined runs 1-6, i.e. for the whole experimental work	0.142	-0.104

Results of all the samples, i.e. pupils participating in 3 sessions, were used in the Pearson-r value calculation for each run of the experiment)

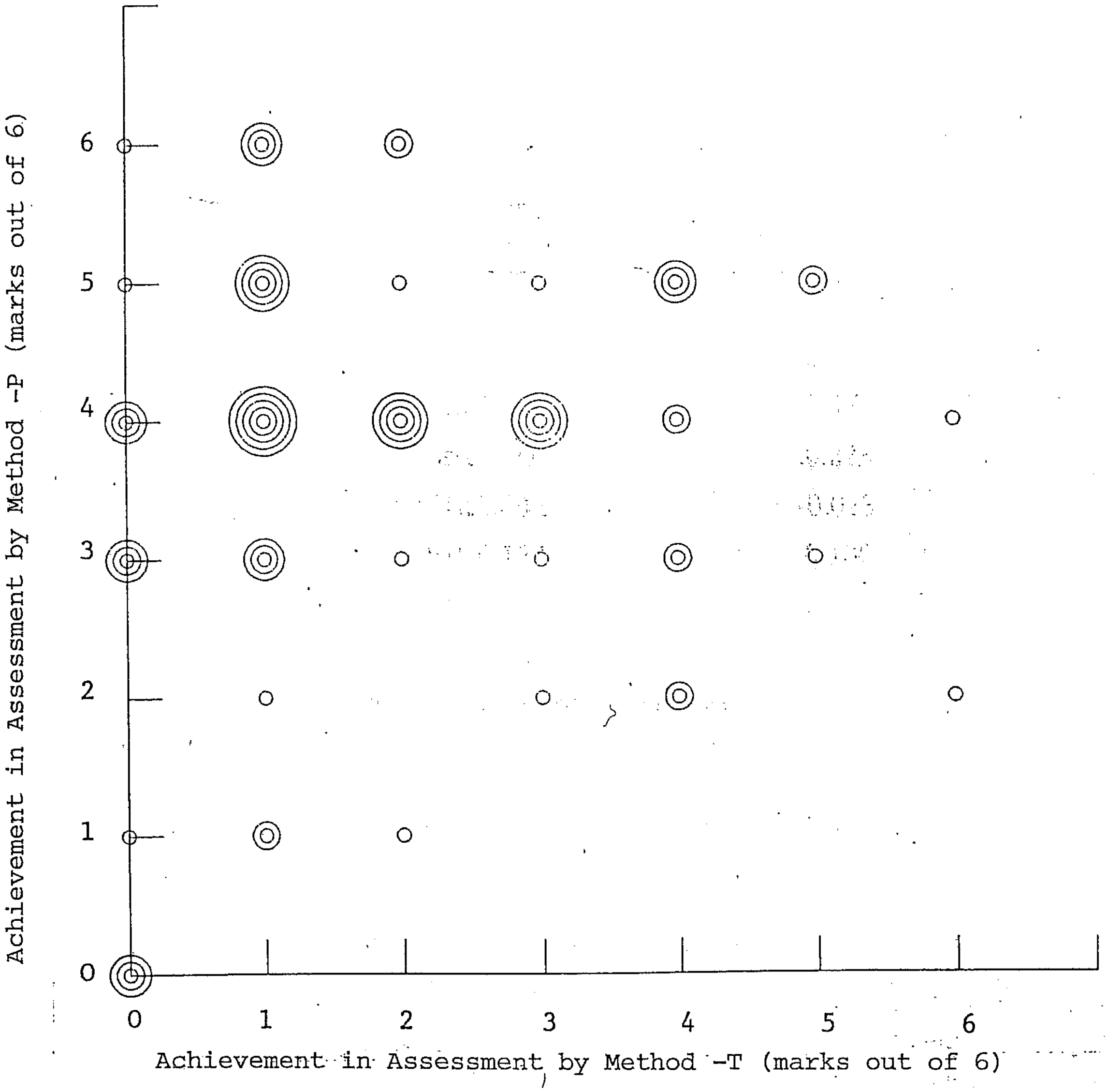

Fig. 2.1 Plots of achievement in assessment (on-the-spot observation) by Method -T for measuring volume by the pupils participating in 3 sessions against their achievement in assessment by Method -P (checking the reported results of the experiment) for carrying out the whole experiment 2.1.

(The data are in Tables 2,6, 2.7 and 2.8)

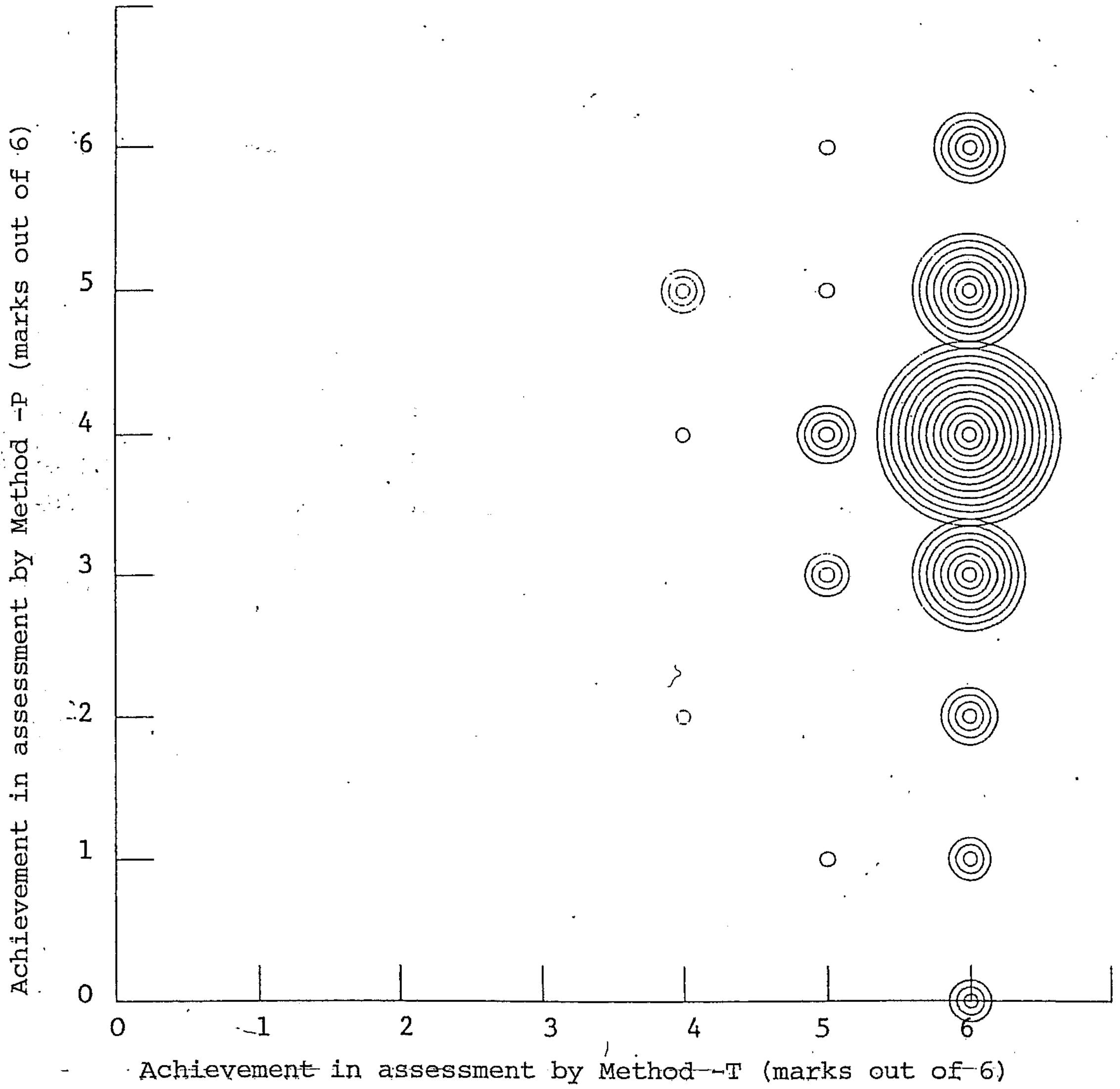

Fig 2.2 Plots of achievement in assessment (on-the-spot observation) by Method -T for pouring out liquid by the pupils participating in 3 sessions against their achievement in assessment by Method -P (checking the reported results of the experiment) for carrying out the whole experiment 2.1.

(The data are in Tables 2.6, 2.7 and 2.8)